附录 C 凸优化基础

无论是最小二乘估计、极大似然估计还是广义矩估计，最终要解决的都是一个最优化问题。凸优化是最优化模型中较容易求解的一类优化问题，线性规划和二次规划等常见的优化问题都是凸优化的特例。在 20 世纪六七十年代，非线性优化问题得到学者的广泛关注，一些找到局部最优解的方法被提出，但是这些方法在寻找全局最优解问题上遇到困难。所以在当时形成了线性优化问题是容易处理的，而非线性优化问题通常较难处理的观念。到了 20 世纪八十年代后期，前苏联学者发现使最优化问题容易处理的关键性质并非是线性，而是凸性。现如今，求解凸优化问题的算法已经相对比较成熟，许多软件都提供了专门求解凸优化问题的程序包，例如，MATLAB 的 CVX、R 的 CVXR 以及 Python 的 CVXPY。对于大多数应用最优化技术的金融计量学者来说，通常的任务是将实践中遇到的具体优化问题通过某种方式转化为凸优化问题，再利用上述软件对问题进行求解。本附录对书中使用到的凸优化知识进行简要地介绍，想进一步了解相关知识的读者可以参考 Calafiore et al. (2014) 或刘浩洋，户将，李勇锋，文再文 (2020)。

C.1 凸集与凸函数

令 C 是 $\mathbb{R}^n$ 的子集，如果连接 C 中任意两点的线段仍然在 C 中，即，对任意的 $\boldsymbol{x}_1, \boldsymbol{x}_2 \in C$ 以及满足 $0 \leqslant \theta \leqslant 1$ 的 θ, 有：

$$\theta \boldsymbol{x}_1 + (1-\theta)\boldsymbol{x}_2 \in C,$$

则称 C 为凸集。图 C.1 中的后两幅图都是非凸集合的例子。

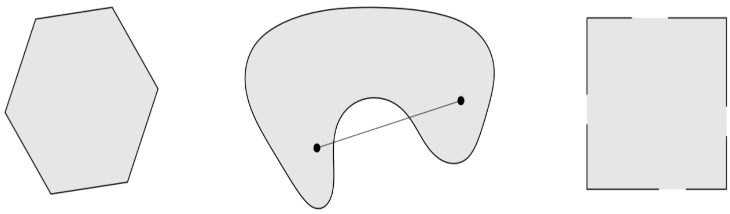

图 C.1: 凸集与非凸集合的例子（图片来源：Boyd et al. (2004) 中图 2.2）

令 f 是由 $\mathbb{R}^n$ 到 $\mathbb{R}$ 的映射，如果 $\mathbf{dom}\, f$ 是凸集，并且，对任意的 $\boldsymbol{x}, \boldsymbol{y} \in \mathbf{dom}\, f$ 以及满足 $0 \leqslant \theta \leqslant 1$ 的 θ，有：

$$f(\theta \boldsymbol{x} + (1-\theta)\boldsymbol{y}) \leqslant \theta f(\boldsymbol{x}) + (1-\theta) f(\boldsymbol{y}), \tag{C.1}$$

则称 f 为凸函数。如果 $-f$ 为凸函数，则称 f 为凹函数。

若对任意的 $\boldsymbol{x}, \boldsymbol{y} \in \mathbf{dom}\, f$ 且 $\boldsymbol{x} \neq \boldsymbol{y}$，$0 < \theta < 1$，有：

$$f(\theta \boldsymbol{x} + (1-\theta)\boldsymbol{y}) < \theta f(\boldsymbol{x}) + (1-\theta) f(\boldsymbol{y}), \tag{C.2}$$

则称 f 为严格凸函数。

图 C.2 给出了一典型的严格凸函数的图像，从图中可以看出，连接严格凸函数曲线上任意两点的线段必然位于这两点之间的函数曲线的上方。

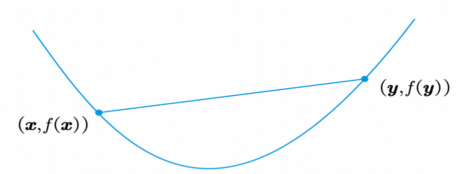

图 C.2: 凸函数

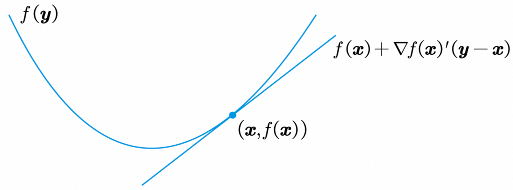

图 C.3: 凸函数的一阶条件

判定一函数是否为凸函数可以通过如下的一阶条件和二阶条件：

定理 C.1. 凸函数判定的一阶条件和二阶条件

设 $f: \mathbb{R}^n \to \mathbb{R}$ 是可微函数（即，对于开集 $\mathbf{dom}\, f$ 内的任意一点 $\boldsymbol{x}$，$\nabla f(\boldsymbol{x})$ 存在），则有：f 是凸函数当且仅当 $\mathbf{dom}\, f$ 是凸集且

$$f(\boldsymbol{y}) \geqslant f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})'(\boldsymbol{y} - \boldsymbol{x}), \quad \forall\, \boldsymbol{x},\, \boldsymbol{y} \in \mathbf{dom}\, f.$$

若进一步假设 f 是二阶可微函数（即，对于开集 $\mathbf{dom}\, f$ 内的任意一点 $\boldsymbol{x}$，二阶海塞 $\nabla^2 f(\boldsymbol{x})$ 存在），则有：f 是凸函数当且仅当 $\mathbf{dom}\, f$ 是凸集且

$$\nabla^2 f(\boldsymbol{x}) \succeq \mathbf{0}, \quad \forall\, \boldsymbol{x} \in \mathbf{dom}\, f.$$

♡

此定理的证明可以参考 Bertsekas (2009) 的命题 1.1.7 和 1.1.10。图 C.3 说明了凸函数一阶条件的几何意义，即，过凸函数上任意一点的切线位于凸函数曲线的下方。此外，从一阶条件可以看出，若 $\nabla f(\boldsymbol{x}) = \mathbf{0}$，则有 $f(\boldsymbol{y}) \geqslant f(\boldsymbol{x})$，即，$f(\boldsymbol{x})$ 为 f 的全局最小值。

令 f 是由 $\mathbb{R}^n$ 到 $\mathbb{R}$ 的映射，f 的 α-下水平集（sublevel set）定义为：

$$C_\alpha = \{\boldsymbol{x} \in \mathbf{dom}\, f \mid f(\boldsymbol{x}) \leqslant \alpha\}.$$

对任意的 α，凸函数的 α-下水平集一定是凸集；逆命题不成立，例如，考虑函数 $-e^x$。f 的上方图（epigraph）定义为：

$$\mathbf{epi}\, f = \{(\boldsymbol{x}, t) \mid \boldsymbol{x} \in \mathbf{dom}\, f, f(\boldsymbol{x}) \leqslant t\} \subseteq \mathbf{R}^{n+1}.$$

f 是凸函数当且仅当它的上方图为凸集。

C.2 凸优化问题

标准的凸优化问题具有如下形式：

$$\begin{aligned} \min \quad & f_0(\boldsymbol{x}) \\ \text{s.t.} \quad & f_i(\boldsymbol{x}) \leqslant 0, \quad i = 1, \ldots, m, \\ & \boldsymbol{a}_j' \boldsymbol{x} = b_j, \quad j = 1, \ldots, p, \end{aligned} \tag{C.3}$$

其中，$f_0, \ldots, f_m$ 是凸函数。我们称 f_0 为目标函数，$f_i(\boldsymbol{x}) \leqslant 0$ 为不等式约束条件，$\boldsymbol{a}_j' \boldsymbol{x} = b_j$ 为等式约束条件。凸优化问题 (C.3) 的定义域定义为：

$$\mathcal{D} = \bigcap_{i=0}^{m} \mathbf{dom}\, f_i.$$

凸优化问题 (C.3) 的可行集（feasible set）定义为：

$$\mathcal{X}=\left\{\boldsymbol{x} \mid f_i(\boldsymbol{x}) \leqslant 0,\ i=1,\dots,m;\ \boldsymbol{a}_j'\boldsymbol{x}=b_j,\ j=1,\dots,p\right\}.$$

容易验证：凸优化问题的可行集一定是凸集。因此，求解凸优化问题是在一个凸集上寻找一个凸目标函数的最小值。对于凸优化问题，任意的局部最优解也是全局最优解（证明请参考 Boyd et al. (2004) 的第 138 页）。

若凸优化问题的目标函数 f_0 可微，则有如下最优条件：（证明请参考 Boyd et al. (2004) 的第 139 页）

定理 C.2. 可微凸优化问题的最优条件

假设凸优化问题 (C.3) 的目标函数 f_0 可微，令 $\mathcal{X}$ 是此问题的可行集，则有：$\boldsymbol{x}$ 是最优解当且仅当 $\boldsymbol{x}\in\mathcal{X}$ 且

$$\nabla f_0(\boldsymbol{x})'(\boldsymbol{y}-\boldsymbol{x}) \geqslant 0, \quad \forall\, \boldsymbol{y}\in\mathcal{X}. \tag{C.4}$$

♡

此定理有如下重要的推论：

- 对于无约束的凸优化问题，最优条件 (C.4) 简化为：

$$\nabla f_0(\boldsymbol{x})=\mathbf{0}. \tag{C.5}$$

- 对于只有等式约束 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}$ 的凸优化问题，由最优条件 (C.4) 可以推出最优解满足：存在 $\boldsymbol{\lambda}\in\mathbb{R}^p$ 使得

$$\nabla f_0(\boldsymbol{x})+\boldsymbol{A}'\boldsymbol{\lambda}=\mathbf{0}. \tag{C.6}$$

  (C.6) 与 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}$ 一起构成了经典的拉格朗日乘数最优条件。

事实上，对于一般的凸优化问题，有如下的 Karush-Kuhn-Tucker (KKT) 最优条件：

定理 C.3. KKT 最优条件

考虑凸优化问题 (C.3)，假设 $f_0,f_1,\dots,f_m$ 可微，$\mathcal{D}$ 的内点集 $\mathbf{int}\,\mathcal{D}$ 非空，并且，存在 $\boldsymbol{x}\in\mathbf{int}\,\mathcal{D}$ 使得：

$$f_i(\boldsymbol{x})<0,\quad i=1,\dots,m;\quad \boldsymbol{a}_j'\boldsymbol{x}=b_j,\quad j=1,\dots,p,\quad \text{（Slater 条件）}$$

则有：$\boldsymbol{x}^\star$ 是 (C.3) 的最优解，当且仅当，存在 m 维向量 $\boldsymbol{\lambda}^\star$ 和 p 维向量 $\boldsymbol{\nu}^\star$ 使得：

$$\begin{aligned}
f_i\left(\boldsymbol{x}^\star\right) \leqslant 0, &\quad i=1,\dots,m\\
\boldsymbol{a}_j'\boldsymbol{x}^\star=b_j, &\quad j=1,\dots,p\\
\lambda_i^\star \geqslant 0, &\quad i=1,\dots,m\\
\lambda_i^\star f_i\left(\boldsymbol{x}^\star\right)=0, &\quad i=1,\dots,m\\
\nabla f_0\left(\boldsymbol{x}^\star\right)+\sum_{i=1}^m \lambda_i^\star \nabla f_i\left(\boldsymbol{x}^\star\right)+\sum_{j=1}^p \nu_j^\star \boldsymbol{a}_j=\mathbf{0}. &
\end{aligned}$$

♡